BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e36513
doi: 10.3897/biss.3.36513

OPEN ACCESS

Conference Abstract

SpOccSum: An easy-to-use Python tool to summarize species occurrence data from material examined lists in taxonomic revisions

Michael Trizna[‡], Torsten Dikow[§]

‡ Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, United States of America
§ Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States of America

Corresponding author: Torsten Dikow (dikowt@si.edu)

Received: 24 May 2019 | Published: 18 Jun 2019

Citation: Trizna M, Dikow T (2019) SpOccSum: An easy-to-use Python tool to summarize species occurrence data from material examined lists in taxonomic revisions. Biodiversity Information Science and Standards 3: e36513. https://doi.org/10.3897/biss.3.36513

Abstract

Taxonomic revisions contain crucial biodiversity data in the material examined sections for each species. In entomology, material examined lists minimally include the collecting locality, date of collection, and the number of specimens of each collection event. Insect species might be represented in taxonomic revisions by only a single specimen or hundreds to thousands of specimens. Furthermore, revisions of insect genera might treat small genera with few species or include tens to hundreds of species. Summarizing data from such large and complex material examined lists and revisions is cumbersome, time-consuming, and prone to errors. However, providing data on the seasonal incidence, abundance, and collecting period of species is an important way to mobilize primary biodiversity data to understand a species's occurrence or rarity. Here, we present SpOccSum (Species Occurrence Summary)—a tool to easily obtain metrics of seasonal incidence from specimen occurrence data in taxonomic revisions. SpOccSum is written in Python (Python Software Foundation 2019) and accessible through the Anaconda Python/R Data Science Platform as a Jupyter Notebook (Kluyver et al. 2016). The tool takes a simple list of specimen data containing species name, locality, date of collection (preferably

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

separated by day, month, and year), and number of specimens in CSV format and generates a series of tables and graphs summarizing:

1. number of specimens per species,
2. number of specimens collected per month,
3. number of unique collection events, as well as
4. earliest, and
5. most recent collecting year of each species.

The results can be exported as graphics or as csv-formatted tables and can easily be included in manuscripts for publication. An example of an early version of the summary produced by SpOccSum can be viewed in Tables 1, 2 from Markee and Dikow (2018). To accommodate seasonality in the Northern and Southern Hemispheres, users can choose to start the data display with either January or July. When geographic coordinates are available and species have widespread distributions spanning, for example, the equator, the user can itemize particular regions such as North of Tropic of Cancer (23.5°N), Tropic of Cancer to the Equator, Equator to Tropic of Capricorn, and South of Tropic of Capricorn (23.5°S). Other features currently in development include the ability to produce distribution maps from the provided data (when geographic coordinates are included) and the option to export specimen occurrence data as a Darwin-Core Archive ready for upload to the Global Biodiversity Information Facility (GBIF).

Table 1.

Collection event summary for *Microphontes* species (Insecta: Diptera: Asilidae) from Markee and Dikow (2018).

Species	**# Specimens**	**# Collection events**	**Earliest collection**	**Most recent collection**
M. ericfisheri	1	1	2015	2015
M. gaiophanes	17	1	2017	2017
M. jasonlondti	4	3	1986	1998
M. kryphios	4	2	1990	2002
M. megoura	9	1	1936	1936
M. safra	5	3	1974	2012
M. whittingtoni	3	2	1990	2008
Microphontes sp.	1	1	1999	1999
summary	44	14	1936	2017

Table 2.

Seasonal incidence of *Microphontes* species (Insecta: Diptera: Asilidae) through number of specimens collected and unique collection events in each month (data given as # specimens/# collecting events when more than one specimen has been collected). Months abbreviated starting with July. From Markee and Dikow (2018).

Species	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
M. ericfisheri	-	-	-	-	-	1	-	-	-	-	-	-
M. gaiophanes	-	-	17/1	-	-	-	-	-	-	-	-	-
M. jasonlondti	-	-	-	-	4/3	-	-	-	-	-	-	-
M. kryphios	-	-	-	-	4/2	-	-	-	-	-	-	-
M. megoura	-	-	-	-	9/1	-	-	-	-	-	-	-
M. safra	-	-	-	-	-	-	-	5/3	-	-	-	-
M. whittingtoni	-	-	-	-	3/2	-	-	-	-	-	-	-
Microphontes sp.	-	-	-	-	1	-	-	-	-	-	-	-
total	-	-	17/1	-	21/9	1	-	5/3	-	-	-	-

Keywords

biodiversity data, species occurrence, seasonal incidence, Python, Jupyter Notebook

Presenting author

Mike Trizna and Torsten Dikow

Presented at

Biodiversity_Next 2019

References

- Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C (2016) Jupyter Notebooks -- a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B (Eds) Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press BV, Amsterdam, 87-90 pp. URL: http://orca.cf.ac.uk/id/eprint/108581

- Markee A, Dikow T (2018) Taxonomic revision of the assassin-fly genus Microphontes Londt, 1994 (Insecta, Diptera, Asilidae). African Invertebrates 59 (2): 195-237. https://doi.org/10.3897/AfrInvertebr.59.30684
- Python Software Foundation (2019) Python Language Reference, version 3.7. https://www.python.org/. Accessed on: 2019-4-21.